*For the sigmod function*

*We notice that the derivative* =

=> x)

=>

NLL(ω)=

=>

=>

=>

↓

***1.2)a)*** *Likelihood* =>

=>

=>

=> *Log likelihood =*

***b)***

***c)****If xt is very sparse the time complexity for the update rule is linear.*

***d)****With very large we will have wild oscillations around the optimum and we could skip the optimal solution.*

***e)***

***f)****The update from the gradient of the log likelihood only touches the dimensions for which inspect vector xj is non zero. Thus, if xi is sparse => The gradient update for log likelihood is sparse. In case of regularisation, the gradient is non-zero at every non-zero parameter dimension.*

*If we assume*

=>

*=> We can deal with regularisation updates by batching them up. Thus instead of doing a matrix multiplication, they can be computed just before they are needed in a lazy manner.*

|  |  |  |
| --- | --- | --- |
| Metric | Deceased patients | Alive patients |
| Event Count  1. Average Event Count  2. Max Event Count  3. Min Event Count | 1029.059  16829  2 | 682.6474  12627  1 |
| Encounter Count  1. Average Encounter Count  2. Max Encounter Count  3. Min Encounter Count | 24.861  375  1 | 18.6694  391  1 |
| Record Length  1. Average Record Length  2. Max Record Length  3. Min Record Length | 151.397  2601  0 | 194.6541  3103  0 |
| Common Diagnosis | DIAG320128  DIAG319835  DIAG313217  DIAG197320  DIAG132797 | DIAG320128  DIAG319835  DIAG317576  DIAG42872402  DIAG313217 |
| Common Laboratory Test | LAB3009542  LAB3023103  LAB3000963  LAB3018572  LAB3016723 | LAB3009542  LAB3000963  LAB3023103  LAB3018572  LAB3007461 |
| Common Medication | DRUG19095164  DRUG43012825  DRUG19049105  DRUG956874  DRUG19122121 | DRUG19095164  DRUG43012825  DRUG19049105  DRUG19122121  DRUG956874 |